Request to Archive
With The National Centers for Environmental Information
For ESRL Global Monitoring Division
Provided by NOAA/OAR/ESRL

2015-02-24

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Ken Masarie
NOAA/OAR/ESRL
Research Physicist
303-497-6270
kenneth.masarie@noaa.gov

2. Name the organization or group responsible for creating the dataset.

NOAA ESRL Global Monitoring Division (GMD)

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

The GMD Carbon Cycle Group makes ongoing measurements from a variety of sampling platforms (e.g., ships, aircraft, towers, and surface). Several sampling programs deploy instruments in situ and make measurements at high frequency and in real time. Other sampling programs are based on weekly air samples collected in glass flasks from large (up to 50+ locations) observing networks and shipped to Boulder where they are analysed at NOAA and the University of Colorado for as many as 50 trace gas species and isotopes. We are interested in archiving these long-term, ongoing measurements from the GMD Carbon Cycle Group. We have identified 50 data collections for archive. Each collection is for a single trace gas species. Measurements of a single trace gas species from discrete (weekly) samples collected at many sites of a large network will constitute a single collection. For example, measurements of CO2 from our surface flask network (50+ different sampling locations) will be a single archive collection. Measurements of CH4 from the same network will be another single archive collection. Measurements derived from our in situ sampling programs will each be a single collection. For example, the Mauna Loa atmospheric CO2 record will be a single collection. The Mauna Loa CH4 atmospheric record will be another single archive collection.

Collections will be updated annually. An annual update of a collection will include all data from the previous archive plus an additional year of data. The following are best estimates of data volume and archive rate for each type of data collection. These estimate are based on measurements of CO2. Data volume and rates for all other trace gases will be considerably smaller than those for CO2. Thus these estimates are an upper limit.

+ surface-insitu measurement collection (e.g., MLO from beginning of record through 2013) : 250 MB; 20 MB per annual update
+ tower-insitu measurement collection (e.g., LEF from beginning of record through 2013) : 350 MB; 20 MB per annual update
+ surface-flask network: (beginning of record through 2013) : 650 MB; 10 MB per annual update
+ aircraft-pfp network: (beginning of record through 2013) : 225 MB; 5 MB per annual update
+ surface-pfp network: (beginning of record through 2013): 100 MB; 2 MB per annual update

Based on the above estimates...

CO2: 4.6 GB initial archive (277 MB annual updates) based on 4 surface in situ collections, 9 tower in situ collections, surface-flask, surface-pfp, and aircraft-pfp networks

CH4: 1.8 GB initial archive (80 MB annual update) based on 2 surface in situ collections, 1 tower in situ collection, surface-flask, surface-pfp, and aircraft-pfp networks

CO: 4.2 GB initial archive (230 MB annual update) based on 2 surface in situ collections, 9 tower in situ collections, surface-flask, surface-pfp, and aircraft-pfp networks

d13C (CO2), d18O (CO2), d13C (CH4), d14C (CO2): 975 MB initial archive (34 MB annual update) based on 4 surface-flask, 4 surface-pfp, and 3 aircraft-pfp networks.

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)
From 1973-01-01 to 2013-12-31

5. Edition or version number(s) of the dataset:
N/A

6. Approximate date when the dataset was or will be released to the public:
2015

7. Who are the expected users of the archived data? How will the archived data be used?
GMD carbon cycle data used by international research community, policy makers, educators, students, and laypeople. Archived data collections will be distributed by GMD only.

8. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?
Evaluations based on feedback from data users and data providers. The WMO World Data Center for Greenhouse Gases (WDCGG) also provides regular feedback on annual updates.

9. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?
This is the overall request from ESRL/GMD to archive data at NCDC. It will include multiple collections of data that are related to one another. This request specifically identifies a subset of measurements made by GMD Carbon Cycle Group.

10. List the input datasets and ancillary information used to produce the data.
Comprehensive metadata is included with each data collection described in item 3 above.

11. List web pages and other links that provide information on the data.
Program and measurements described at www.esrl.noaa.gov/gmd/ccgg/. Each data set includes extensive metadata describing the source of the data, providers, calibration scale, comparison activity, and selection strategy. Each data set also includes metadata describing the product as a whole including Data Usage Policy, Required Citation, and Creation information. We will send metadata that follows NCDC's implementation of ISO standards for metadata.

12. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.

1. These products are currently available for download from www.esrl.noaa.gov/gmd/ in various formats. We will send metadata that follows NCDC's implementation of ISO standards for metadata.
2. Additional summary files include e-mail list of data providers, data summary, provider summary, and citation. Also included is a configuration (rc) file describing, for each measurement project, how which source data are used and how they are processed.

13. Indicate the data file format(s).

1. netCDF-4
2. Ascii Text

14. Are the data files compressed?

gzip

15. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).

It is possible to have several data sets at the same location but derived using data from differing sampling strategies (e.g., surface-flask, surface-insitu, aircraft-pfp) and differing selection strategies. The file naming scheme assures uniqueness. In situ data collections will be available by individual files and as a single compressed file of the collection. Likewise, surface flask data for a given trace gas species will be provided as individual files are a single compressed file of the collection.

Examples...

co2_lef_tower-insitu_1_ccgg_HourlyData.nc
co_lef_tower-insitu_1_ccgg_HourlyData.nc
ch4_brw_surface-insitu_1_ccgg_MonthlyData.nc
co2_brw_surface-flask_1_ccgg_AllData.nc
co2_uum_surface-flask_1_ccgg_AllData.nc

16. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?

Four collections used for the NCDC/GMD pilot may be found at ftp://aftp.cmdl.noaa.gov/data/greenhouse_gases/co2/in-situ/surface/. Many of the other data collections mentioned in Item 3 may be found at ftp://aftp.cmdl.noaa.gov/data/trace_gases/. These other data collections will be reformatted to be consistent with the initial pilot collections.

17. What is the total data volume to be submitted?

Historic Data: all historic data or data submitted as a completed collection.

Total Data Volume: 11GB
Number of Data Files: 800

Continuous Data: data volume rate for a continuous data production.

Total Data Volume Rate: 1.5GB per Year
Data File Frequency: 800 per Year
Data Production Start:

18. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

Collections are updated annually or as needed to include updates to existing datasets and the addition of new datasets if available. Updates may include revised values for previously included measurements.

19. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location: Boulder, Colorado
System Name: masarie.cmdl.noaa.gov
System Owner: NOAA ESRL GMD
Additional Information:

20. What are the possible methods for submitting the data to NCEI? Select all that apply.

1. SFTP PUSH

21. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

1. No web access

22. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

Constraint Type	Description
Access	Product will not be distributed from data center. Product discovery metadata will be available from data center. Further discussion with NCDC and GMD required.

23. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.

GMD intends to archive all GMD data with NCDC. In addition to improving access to these critical climate data records, GMD is will want to mint DOIs to their data collections.

24. Are the data archived at another facility or are there plans to do so? Please explain.

All GMD data are currently archived at The World Meteorological Organization (WMO) World Data Centers (WDC). There are 6 WMO WDCs: Ozone, Solar Radiation, GHG, Aerosols, Precipitation Chemistry, and Remote Sensing of Atmosphere. GMD data are archived in the appropriate WMO Data Centers. Per prior requests with NCDC it is still vital to archive at a NOAA Data Center.

25. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?

SLA between ESRL/GMD and NCEI (NCDC) is in the works.

26. Do you have a data management plan for your data?

No

27. Have funds been allocated to archive the data at NCEI?

GMD intends to provide funds to NCDC for archiving collections identified in Item 3 above.

28. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.

N/A

29. Is there a desired deadline for NCEI to archive and provide access to the data?

Archive by: 2015-12-31

Accessible by:

30. Add any other pertinent information for this request.

None